Phylogenomic inference in extremis

Lam, Vivienne K. Y.; Darby, Hayley; Merckx, Vincent S. F. T.; Lim, Gwynne; Yukawa, Tomohisa; Neubig, Kurt M.; Abbott, J. Richard; Beatty, Gemma E.; Provan, Jim; Gomez, Marybel Soto; Graham, Sean W.

Published in:
American Journal of Botany

DOI:
10.1002/ajb2.1070

Publication date:
2018

Citation for published version (APA):
Lam, V. K. Y., Darby, H., Merckx, V. S. F. T., Lim, G., Yukawa, T., Neubig, K. M., Abbott, J. R., Beatty, G. E., Provan, J., Gomez, M. S., & Graham, S. W. (2018). Phylogenomic inference *in extremis*: A case study with mycoheterotroph plastomes. *American Journal of Botany*, *105*(3), 480-494. https://doi.org/10.1002/ajb2.1070

American Journal of Botany

RESEARCH ARTICLE

INVITED SPECIAL ARTICLE
For the Special Issue: Using and Navigating the Plant Tree of Life

Phylogenomic inference *in extremis*: A case study with mycoheterotroph plastomes

Vivienne K. Y. Lam[1,2], Hayley Darby[1,2], Vincent S. F. T. Merckx[3], Gwynne Lim[4,5], Tomohisa Yukawa[6], Kurt M. Neubig[7], J. Richard Abbott[8], Gemma E. Beatty[9], Jim Provan[9], Marybel Soto Gomez[1,2], and Sean W. Graham[1,2,10]

Manuscript received 10 October 2017; revision accepted 2 February 2018.

[1] Department of Botany, University of British Columbia, 6270 University Boulevard, Vancouver, British Columbia V6T 1Z4, Canada

[2] UBC Botanical Garden & Centre for Plant Research, University of British Columbia, 6804 Marine Drive SW, Vancouver, British Columbia V6T 1Z4, Canada

[3] Naturalis Biodiversity Center, Vondellaan 55, 2332 AA, Leiden, The Netherlands

[4] The New York Botanical Garden, Pfizer Plant Research Laboratory, 2900 Southern Boulevard, Bronx, NY 10458, USA

[5] L. H. Bailey Hortorium, Section of Plant Biology, Cornell University, 412 Mann Library Building, Ithaca, NY 14853, USA

[6] Tsukuba Botanical Garden, National Science Museum, Tsukuba, Japan

[7] Department of Plant Biology, Southern Illinois University, Carbondale, IL 62901, USA

[8] Missouri Botanical Garden, P.O. Box 299, St. Louis, MO 63166-0299, USA

[9] Institute of Biological, Environmental& Rural Sciences (IBERS), Aberystwyth University, Aberystwyth, Ceredigion SY23 3EE, UK

[10] Author for correspondence (e-mail: swgraham@interchange.ubc.ca)

Citation: Lam, V. K. Y., H. Darby, V. S. F. T. Merckx, G. Lim, T. Yukawa, K. M. Neubig, J. R. Abbott, et al. 2018. Phylogenomic inference *in extremis*: A case study with mycoheterotroph plastomes. *American Journal of Botany* 105(3): 1–15.

doi:10.1002/ajb2.1070

PREMISE OF THE STUDY: Phylogenomic studies employing large numbers of genes, including those based on plastid genomes (plastomes), are becoming common. Nonphotosynthetic plants such as mycoheterotrophs (which rely on root-associated fungi for essential nutrients, including carbon) tend to have highly elevated rates of plastome evolution, substantial genome reduction, or both. Mycoheterotroph plastomes therefore provide excellent test cases for investigating how extreme conditions impact phylogenomic inference.

METHODS: We used parsimony and likelihood analysis of protein-coding gene sets from published and newly completed plastomes to infer the phylogenetic placement of taxa from the 10 angiosperm families in which mycoheterotrophy evolved.

KEY RESULTS: Despite multiple very long branches that reflect elevated substitution rates, and frequently patchy gene recovery due to genome reduction, inferred phylogenetic placements of most mycoheterotrophic lineages in DNA-based likelihood analyses are both well supported and congruent with other studies. Amino-acid-based likelihood placements are broadly consistent with DNA-based inferences, but extremely rate-elevated taxa can have unexpected placements—albeit with weak support. In contrast, parsimony analysis is strongly misled by long-branch attraction among many distantly related mycoheterotrophic monocots.

CONCLUSIONS: Mycoheterotrophic plastomes provide challenging cases for phylogenomic inference, as substitutional rates can be elevated and genome reduction can lead to sparse gene recovery. Nonetheless, diverse likelihood frameworks provide generally well-supported and mutually concordant phylogenetic placements of mycoheterotrophs, consistent with recent phylogenetic studies and angiosperm-wide classifications. Previous predictions of parallel photosynthesis loss within families are supported for Burmanniaceae, Ericaceae, Gentianaceae, and Orchidaceae. Burmanniaceae and Thismiaceae should not be combined as a single family in Dioscoreales.

KEY WORDS Corsiaceae; incomplete multigene alignments; Iridaceae; long-branch attraction; mycoheterotrophy; orchids; Petrosaviaceae; photosynthesis loss; Polygalaceae; Triuridaceae.

Improvements in sequencing technologies in the last decade have made it feasible to assemble whole plastid genomes (plastomes) quite cheaply and rapidly (Goodwin et al., 2016), so that over 2000 plastomes are now available on GenBank (NCBI Organellar Genomes; December 2017). Similar rapid advances with the production of transcriptomes and whole genomes for hundreds or even

thousands of plant species are taking place, or are anticipated soon (Wickett et al., 2014; Normille, 2017). The resulting massive influxes of data will undoubtedly contribute to more refined pictures of plant relationships. For example, studies based on whole-plastid genomes of photosynthetic plants have helped to resolve several problematic relationships at broad and recent levels of plant phylogeny, confirmed earlier phylogenetic results based on few-gene data sets, and affirmed the recalcitrance of several hard-to-resolve relationships (e.g., Givnish et al., 2010; Ruhfel et al., 2014; Barrett et al., 2016; Ross et al., 2016; Fishbein et al., 2018 and Gitzendanner et al., 2018). However, it will take a while to fully integrate some lineages into the plant tree of life—for example, those with highly modified, rapidly evolving, or difficult-to-recover genomes. The latter phenomena are observed in the plastomes of many heterotrophic (nonphotosynthetic) plants (e.g., Lam et al., 2015; Mennes et al., 2015a; Bellot and Renner, 2016; Bellot et al., 2016; Naumann et al., 2016; Wicke et al., 2016).

Highly elevated substitution rates may lead to long-branch attraction, resulting in phylogenetic misinference (Felsenstein, 1978; Hendy and Penny, 1989). Elevated substitution rates, sometimes extreme (e.g., Naumann et al., 2016), are typical of many heterotrophic plant lineages and have been shown to introduce systematic error in phylogenetic inference (e.g., Barkman et al., 2004; Nickrent et al., 2004; Merckx et al., 2009; Lam et al., 2015). This phenomenon is not limited to heterotrophic plants, as several photosynthetic lineages also have notably elevated rates, including the hard-to-place Gnetales (e.g., Sanderson et al., 2000; Graham and Iles, 2009) and Poales in monocots (e.g., Givnish et al., 2010). Heterotrophic lineages can also display considerable gene loss, leading to patchily populated multigene alignments for phylogenomic inference. Gene loss could also contribute to difficulties in phylogenetic inference, although adding taxa with only sparse gene recovery is expected to help resolve relationships at previously uncertain nodes (e.g., Burleigh et al., 2009; Wiens and Tiu, 2012; Jiang et al., 2014). While typical of many heterotroph plastomes, sparse gene recovery may also occur in photosynthetic lineages when representative specimens do not sequence well (e.g., for DNAs recovered from degraded material, such as older herbarium specimens; Statts et al., 2013). Nonetheless, rate elevation and patchy recovery are expected to be particularly severe in heterotrophic lineages and may typically co-occur. These plants thus arguably represent the most challenging cases for inclusion in phylogenomic studies.

Heterotrophic plants include holoparasites and full mycoheterotrophs, which respectively derive their nutrition from plant or fungal partners. Fully heterotrophic plants are relatively rare in terms of species numbers, representing less than 1% of land-plant species (Merckx, 2013). However, they have evolved repeatedly across plant phylogeny (at least 11 origins of holoparasitism, and at least 47 origins of full mycoheterotrophy; Merckx et al., 2013c; Nickrent, 2017). There are over 500 known species that are fully mycoheterotrophic, mostly in the angiosperms. A minimum of seven origins of full mycoheterotrophy (representing ~50 species) are known in the core eudicots (Merckx et al., 2013c), where full mycoheterotrophy has evolved independently in three distantly related families (Ericaceae, Gentianaceae, and Polygalaceae). Partial mycoheterotrophs (plants that both photosynthesize and derive some nutrition from fungal partners) are also known in Ericaceae and Gentianaceae (Tedersoo et al., 2007; Zimmer et al., 2007; Hynson et al., 2009; Cameron and Bolin, 2010; Merckx et al., 2013c).

Monocots appear to be particularly susceptible to evolving full mycoheterotrophy, with the majority of known mycoheterotrophic species (91%) and origins (83%) found in seven monocot families containing fully mycoheterotrophic taxa (Burmanniaceae, Corsiaceae, Iridaceae, Orchidaceae, Petrosaviaceae, Thismiaceae, and Triuridaceae; Leake, 1994; Imhof, 2010; Merckx et al., 2013c). All orchids are initially mycoheterotrophic during seedling establishment, but some species are also partially or fully mycoheterotrophic at maturity: there have been an estimated 30 independent losses of photosynthesis in orchids alone (Merckx and Freudenstein, 2010), yielding perhaps 235 fully mycoheterotrophic species (Bernard, 1909; Leake, 1994; Rasmussen, 1995; Merckx, 2013; Merckx et al., 2013a, c). Burmanniaceae (Dioscoreales) also have multiple independent losses of photosynthesis with eight genera and 96 species that likely range from partial to full mycoheterotrophs, and at least eight losses of photosynthesis (Merckx et al., 2006, 2013c; Bolin et al., 2017).

Here we focus on the phylogenetic placement of mycoheterotrophic plants using plastid genome-scale data sets. Their sometimes extreme morphological modifications and the loss of the plastid genes used for phylogenetic inferences pertinent to angiosperm-wide classification have contributed to uncertainty about the phylogenetic placement of multiple mycoheterotrophic lineages until very recently (summarized by Lam et al., 2016). However, a preliminary phylogenetic study of monocot mycoheterotrophs using a few plastid genes commonly retained in heterotrophic plants (*accD*, *clpP*, *matK*) demonstrated that they retain sufficiently useful phylogenetic signal to help place them in monocot-wide phylogenetic history—even when very rapidly evolving or with only one or two genes recovered, at least when analyzed using model-based methods (Lam et al., 2016). Their pilot study was based on a few (one to three) genes recovered using Sanger sequencing, and the inferred phylogenetic placements of mycoheterotrophic families were poorly supported in several cases (Burmanniaceae, Thismiaceae, Triuridaceae). Whole-plastid genomes (or more specifically, extracted gene sets representing most of the plastid-encoded protein-coding loci) are now being used to place some mycoheterotrophic lineages, such as *Petrosavia stellaris* as the sister group of *Japonolirion* in Petrosaviaceae, Petrosaviales (Logacheva et al., 2014), Corsiaceae as the sister group of Campynemataceae in Liliales (Mennes et al., 2015a; Bodin et al., 2016; Givnish et al., 2016), and Triuridaceae as the sister group of Cyclanthaceae and Pandanaceae in Pandanales (Lam et al., 2015). However, an integrative large-scale phylogenomic analysis that includes multiple independent mycoheterotrophic lineages has not been attempted before, as the plastid genomes of many taxa have not been sampled to date (mycoheterotrophic monocots in Burmanniaceae, Iridaceae; mycoheterotrophic eudicots in Gentianaceae, Polygalaceae) or included previously in analysis (the monocot family Thismiaceae; Lim et al. 2016).

Here we simultaneously analyze previously published (Delannoy et al., 2011; Logacheva et al., 2011, 2014, 2016; Barrett and Davis, 2012; Barrett et al., 2014; Lam et al., 2015; Mennes et al., 2015a; Schelkunov et al., 2015; Feng et al., 2016; Gruzdev et al., 2016; Lim et al., 2016) and newly produced plastid genomes from all 10 of the angiosperm families that include mycoheterotrophs, to explore whether phylogenomic inference permits well-supported resolution of mycoheterotroph relationships that accords with our current broad understanding of plant phylogeny. In addition, our study provides a broad-scale test for how parsimony and model-based methods perform when there are extreme cases of rate elevation or genome reduction in multiple parallel lineages, here in the context of a well-sampled phylogeny that otherwise comprises photosynthetic angiosperms. Rapid molecular evolution and limited gene

sampling are also important phenomena to consider in phylogenomic analysis of holoparasitic plants, as they can also exhibit substantial rate elevation and gene loss (e.g., Bellot et al., 2016; Graham et al., 2017), and are relevant to any photosynthetic lineages with elevated substitution rates and/or limited gene samplings.

MATERIALS AND METHODS

Taxon and gene sampling

We generated new plastid sequences representing retrievable protein-coding genes for 18 fully or partially mycoheterotrophic taxa and five green relatives (Appendices S1–S3, see Supplemental Data with this article), and added these to a published angiosperm plastome matrix (Lam et al., 2015), along with taxa included by Barrett et al. (2014), Givnish et al. (2010, 2015), Ruhfel et al. (2014), Mennes et al. (2015a), and others (Appendix S2 and references cited there). For most mycoheterotrophs, we are confident that we retrieved the full complement of retained protein-coding genes because we were able to recover complete plastid genomes (i.e., new taxa noted here with genome lengths; Appendix S3); these full genomes will be presented elsewhere. The full sampling includes major lineages within the monocots, eudicots, magnoliids and ANA-grade orders (Amborellales, Nymphaeales, and Austrobaileyales). We also aimed for denser sampling in families and orders thought to contain mycoheterotrophic taxa. We did not sample a comprehensive range of parasitic plants as our sampling focus is on mycoheterotrophic plants, although several parasitic eudicot lineages were included (*Cuscuta* in Convolvulaceae; multiple Orobanchaceae). In several cases, available alignments omitted sections of genes, and so we reassembled gene sets for these from genomes available on GenBank.

DNA isolation and library preparation

We isolated DNA following a modified CTAB protocol (Doyle and Doyle, 1987; Rai et al., 2003), and produced sequencing libraries using several library preparation kits. We used the Bioo Nextflex DNA sequencing kit (Bioo Scientific Corp., Austin, TX, USA) for samples with ample starting DNA concentration, and KAPA LTP Library Preparation (KAPA Biosystems, Boston, MA, USA) and NuGEN Ovation Ultralow Library System (NuGEN Technologies, San Carlos, CA, USA) kits for samples with <10 ng of starting DNA. We sheared DNAs to 400-bp fragments for library preparation on a Covaris S220 sonicator (Covaris, Woburn, MA, USA), and size-selected for finished libraries with 550–650-bp fragments. We quantified all libraries using Qubit fluorometry (ThermoFisher Scientific, Waltham, MA, USA), verified library fragment sizes using a Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA), and measured final library concentrations using qPCR on an iQ5 real-time system (Illumina DNA standard kit, KAPA Biosystems, Boston, MA, USA; Bio-Rad Laboratories, Hercules, CA, USA). The libraries were multiplexed (10–20 samples per lane) and sequenced on a HiSeq 2000 platform (Illumina, San Diego, CA, USA) as 100-bp paired-end reads.

De novo contig assembly, plastid gene retrieval

Illumina reads were processed using CASAVA 1.8.2. (Illumina) to sort the multiplexed data by taxon. We performed de novo assembly for each taxon using CLC Genomics Workbench 6.5.1 (CLC bio, Aarhus, Denmark) with default settings, selecting for all contigs >500 bp long with >20× coverage. We then used a custom Perl script (https://github.com/daisieh/phylogenomics/blob/master/filtering/filter_cp.pl) to BLAST contigs against a local database (Altschul et al., 1990) to retrieve plastid contigs, using *Dioscorea elephantipes* (GenBank accession NC_009601.1), *Asclepias syriaca* (NC_022432.1), *Glycine max* (NC_007942.1), and *Arbutus unedo* (JQ067650) as reference taxa. We used Sanger sequencing to connect contigs into full plastid genomes. We annotated plastid genes using DOGMA (Wyman et al., 2004) and manually inspected for gene and exon boundaries using Sequencher v.4.8 (Gene Godes, Ann Arbor, MI, USA) and the species noted above as reference taxa.

Multigene alignments

We extracted genes from plastome assemblies and aligned them individually, excluding intron regions, in Se-Al 2.0a11 (Rambaut, 2002). We omitted putative pseudogenes in mycoheterotrophs (genes with interrupted reading frames due to internal stop codons, resulting from nontriplet indels or substitutions), but included genes with minor reading frame reductions at the start or end of genes that were otherwise in-frame. We aligned each gene individually following Graham et al. (2000) and Saarela and Graham (2010), staggering hard-to-align regions; missing sequences were represented as blanks in individual alignments. The photosynthetic taxa generally are represented by 77 protein-coding genes as we excluded two protein-coding genes, *ycf1* and *ycf2*, that are difficult to align across angiosperms. The fraction of genes recovered for mycoheterotrophs is as low as 9% of the 77 aligned genes considered for photosynthetic taxa (mean 55%, median 49% retention; Appendix S3). Each gene-based file included 398 terminals. We concatenated gene alignments into a single 137,931-bp matrix (derived from 55,696 bp of original, unaligned plastid sequence data in *Premna microphylla*, for reference). We checked for compilation errors in the final matrix by exporting sequences from the concatenated matrix for each added taxon and used Sequencher to compare these back to the original files. We also translated the concatenated matrix into a 45,977-amino-acid residue matrix. The DNA and amino-acid alignments are publicly available at figshare.com (https://doi.org/10.6084/m9.figshare.5480608).

Phylogenetic inference

We analyzed the concatenated plastome alignment with parsimony and maximum likelihood (ML) approaches, using PAUP* 4.0b10 (Swofford, 2003) for parsimony analysis, and RAxML-HPC v.8 on XSEDE (Stamatakis, 2014) for likelihood analyses of unpartitioned and partitioned DNA and amino-acid data sets. The partitioning scheme for the DNA-based matrix initially comprised 231 partitions derived from the first, second, and third codon positions for each protein-coding gene (gene by codon; G × C). We used PartitionFinder 2 (Lanfear et al., 2016) under the r-clustering algorithm (Lanfear et al., 2014) to group partitions with similar substitution models or model parameters under the Bayesian information criterion (BIC; Appendix S4). We found 55 final partitions for the partitioned DNA-based matrix with GTR+G or GTR+I+G DNA substitution models as the best fit for individual partitions (Appendix S4). We used the GTR+G model for all partitions in partitioned analyses (the "I" parameter for invariant sites may be

accommodated by the gamma parameter, G, Yang, 2006). The optimal model for the unpartitioned DNA-based dataset was GTR+I. Additionally, we used PartitionFinder 2 to find the optimal model for the unpartitioned amino acid matrix (here the LG amino-acid substitution model). For the partitioned amino-acid matrix, we employed a smaller subset of 100 representative taxa to find the best partitioning scheme due to computational constraints. We used PartitionFinder 2 to group initial 77 partitions, one for each gene, using the same strategy as for the DNA data set above, but limited the amino-acid substitution models to those currently implemented in RAxML. We inferred 39 final partitions, and applied the optimal models for each one in tree searches (Appendix S4; JTT and HIVB were the most commonly inferred optimal substitution models). Likelihood searches were performed on the CIPRES portal (Miller et al., 2010), with 20 independent searches for the best tree in each case. For the parsimony analysis, we ran a parsimony search for the shortest trees in PAUP* using tree-bisection-reconnection branch swapping (TBR) and 10 random stepwise addition replicates, holding one tree at each step, and otherwise using default settings. We estimated branch support with bootstrap analysis (Felsenstein, 1985), using 300 replicates, with 10 random addition replicates per bootstrap replicate for the parsimony analysis, and 500 rapid bootstrap replicates for the ML analyses (Stamatakis et al., 2008), the latter with the same DNA or amino-acid substitution models and partitioning schemes used in the searches for the best trees, as outlined above. We considered strongly supported branches to have at least 90% bootstrap support, and poorly supported ones to have less than 70% support, following Zgurski et al. (2008).

RESULTS

Likelihood-based phylogenomic analyses

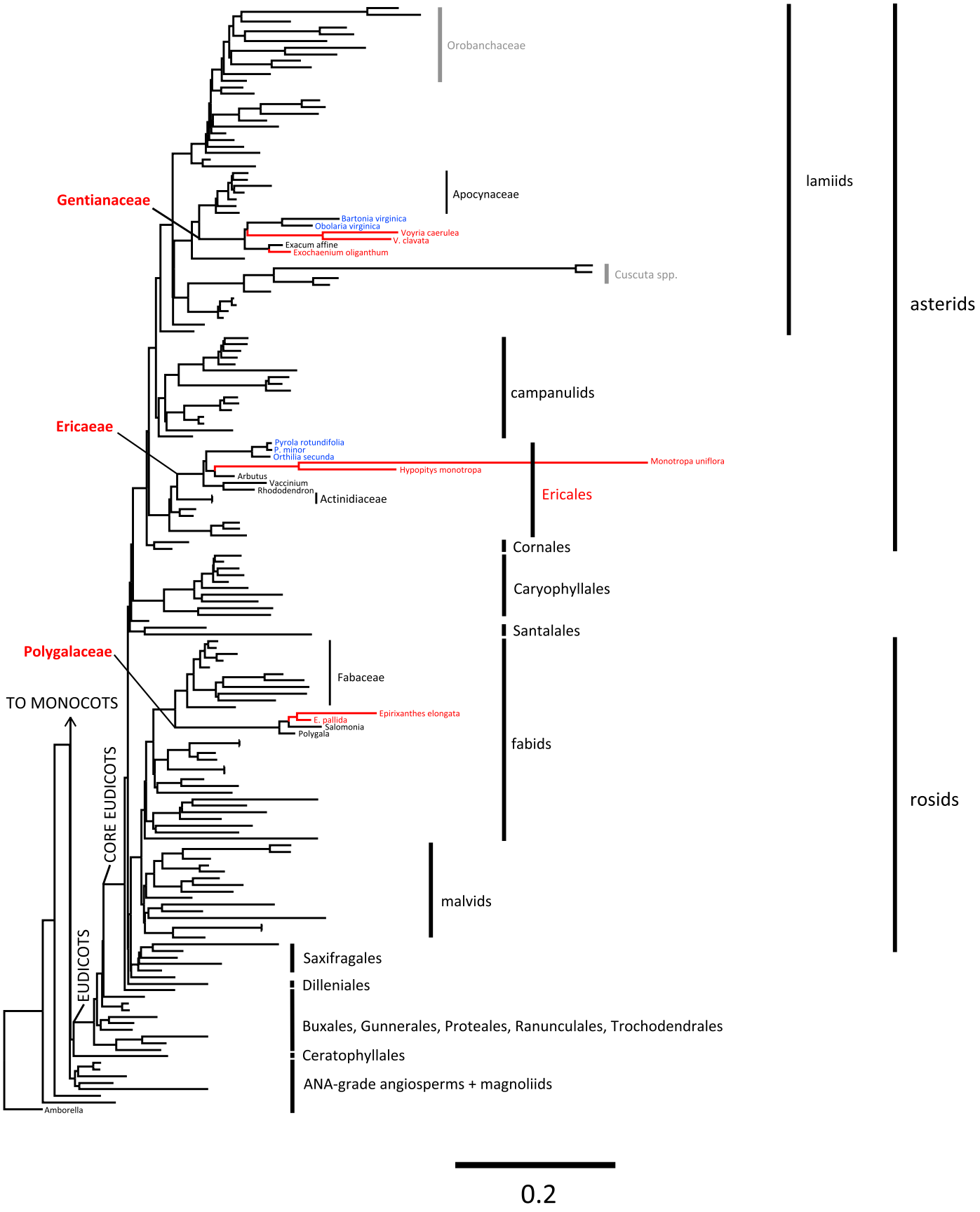
Arachnitis in Corsiaceae, *Epipogium* and *Rhizanthella* (Orchidaceae), fully mycoheterotrophic members of Burmanniaceae, *Thismia* (Thismiaceae), *Sciaphila* (Triuridaceae), and *Monotropa* and *Hypopitys* (Ericaceae) have much longer branches than the other sampled mycoheterotrophs and autotrophic angiosperms (Figs. 1, 2); some of their branch lengths are comparable to or exceed those of holoparasitic *Cuscuta* and Orobanchaceae (note that other holoparasitic lineages with highly elevated rates were not included here). We inferred no major topological differences across the DNA- and amino-acid-based likelihood analyses—apart from weakly supported relationships involving several very long branches in the amino-acid likelihood analyses (Figs. 1–4; Appendices S5–S8). The main topology used for reference here is from the partitioned DNA analysis (Figs. 1, 2), unless noted.

The families Ericaceae, Gentianaceae, and Polygalaceae represent eudicot clades that are distantly related to each other, with consistently strong bootstrap support for the monophyly of each in all likelihood analyses (Fig. 3). The phylogenetic placements of each family were also well supported and consistent across analyses (i.e., among sampled taxa, Ericaceae is the sister group of Actinidiaceae in Ericales, Gentianaceae is the sister group of Apocynaceae in Gentianales, and Polygalaceae is the sister group of Fabaceae in Fabales; Figs. 1, 3; Appendices S5–S8). Within Polygalaceae, we recovered a clade comprising *Epirixanthes* and *Salomonia* as the sister group of *Polygala*, with 99–100% bootstrap support across all likelihood analyses (Fig. 3, Appendices S5–S8). Within Gentianaceae, *Exochaenium* and *Exacum* are inferred to be sister taxa, the partial mycoheterotrophs *Obolaria* and *Bartonia* are sister groups, and the two sampled species of *Voyria* are also sister taxa (Fig. 1), all with 100% bootstrap support across all likelihood analyses (Fig. 3, Appendices S5–S8). The relative arrangement of these three lineages within the family is strongly supported in both DNA-based likelihood analyses, with *Voyria* sister to *Bartonia-Obolaria*, and *Exacum-Exochaenium* then sister to the rest of the family (100% bootstrap support). In contrast, the amino-acid analyses have *Voyria* sister to the rest of the family, although with poor support (Fig. 3). In Ericaceae, the partial mycoheterotrophs *Pyrola* and *Orthilia* are strongly supported as sister taxa among sampled taxa (100% bootstrap support in all analyses; Fig. 3; Appendices S5–S8). Fully mycoheterotrophic *Monotropa* and *Hypopitys* also comprise a strongly supported sister group (99–100% bootstrap support across likelihood analyses), with photosynthetic *Arbutus* as sister to this clade, again with strong support (97–99%). *Vaccinium* and *Rhododendron* are strongly supported sister taxa (100%). There is moderate to strong support for *Vaccinium-Rhododendron* as the sister group of the rest of the family in the DNA-based likelihood analyses (86–99%); this relationship is more poorly supported in the amino-acid-based analyses, and the shortest trees for these analyses instead depict *Vaccinium-Rhododendron* as the sister group of *Orthilia-Pyrola* (Fig. 3; Appendices S5–S8).

The monocot family Petrosaviaceae (= order Petrosaviales) is inferred to be the sister group of all other monocots except *Acorus* and Alismatales, with strong support across all analyses (Figs. 2, 4; Appendices S5–S8). Within Petrosaviaceae, *Petrosavia* is the sister group of the only other genus in the family (photosynthetic *Japonolirion osense*); within *Petrosavia*, *P. stellaris* is sister to a small clade comprising one or two species (*P. sakuraii* and a distinct lineage that may represent a new species) (Fig. 2). All of these inferred relationships are well supported across likelihood analyses (Fig. 4; Appendices S5–S8). *Sciaphila* (Triuridaceae) is inferred to belong to Pandanales with strong support, and is consistently recovered as the sister group of Pandanaceae and Cyclanthaceae with good bootstrap support for this relationship across likelihood analyses (82–91% support; Figs. 2, 4; Appendices S5–S8).

Within Dioscoreales, *Thismia* (Thismiaceae) is inferred to be the sister group of *Tacca* (Taccaceae) in DNA-based analyses (84–87% bootstrap support) (Figs. 2, 4; Appendices S5, S6). In contrast, *Thismia* appears to float close to several other long-branch taxa in

FIGURE 1. Phylogenetic placements of eudicot mycoheterotrophic lineages in the best tree from a "gene by codon" partitioned likelihood analysis of angiosperm plastid genome data (DNA sequence data from 77 protein-coding genes; fewer genes in most mycoheterotrophs, Appendix S3); the tree is continued in Fig. 2. Support values for mycoheterotrophic eudicot families summarized in Fig. 3; terminal labels excluded except in families with fully mycoheterotrophic taxa (see Appendix S5 for full details). Red lineages represent fully mycoheterotrophic taxa (with Dollo parsimony interpretations of photosynthesis loss; re-gain not possible once lost), terminals with blue labels are suspected or known partial mycoheterotrophs (predicted to retain photosynthetic ability), several taxa noted in black are photosynthetic plants (not mycoheterotrophic at maturity); major angiosperm clades and several eudicot parasitic lineages are also noted. The scale bar indicates the estimated number of substitutions per site.


Orobanchaceae
Gentianaceae
Apocynaceae
Bartonia virginica
Obolaria virginica
Voyria caerulea
V. clavata
Exacum affine
Exochaenium oliganthum
Cuscuta spp.
lamiids
asterids
campanulids
Ericaeae
Pyrola rotundifolia
P. minor
Orthilia secunda
Monotropa uniflora
Hypopitys monotropa
Arbutus
Vaccinium
Rhododendron
Actinidiaceae
Ericales
Cornales
Caryophyllales
Santalales
Polygalaceae
Fabaceae
TO MONOCOTS
Epirixanthes elongata
E. pallida
Salomonia
Polygala
fabids
rosids
CORE EUDICOTS
malvids
EUDICOTS
Saxifragales
Dilleniales
Buxales, Gunnerales, Proteales, Ranunculales, Trochodendrales
Ceratophyllales
ANA-grade angiosperms + magnoliids
Amborella
0.2

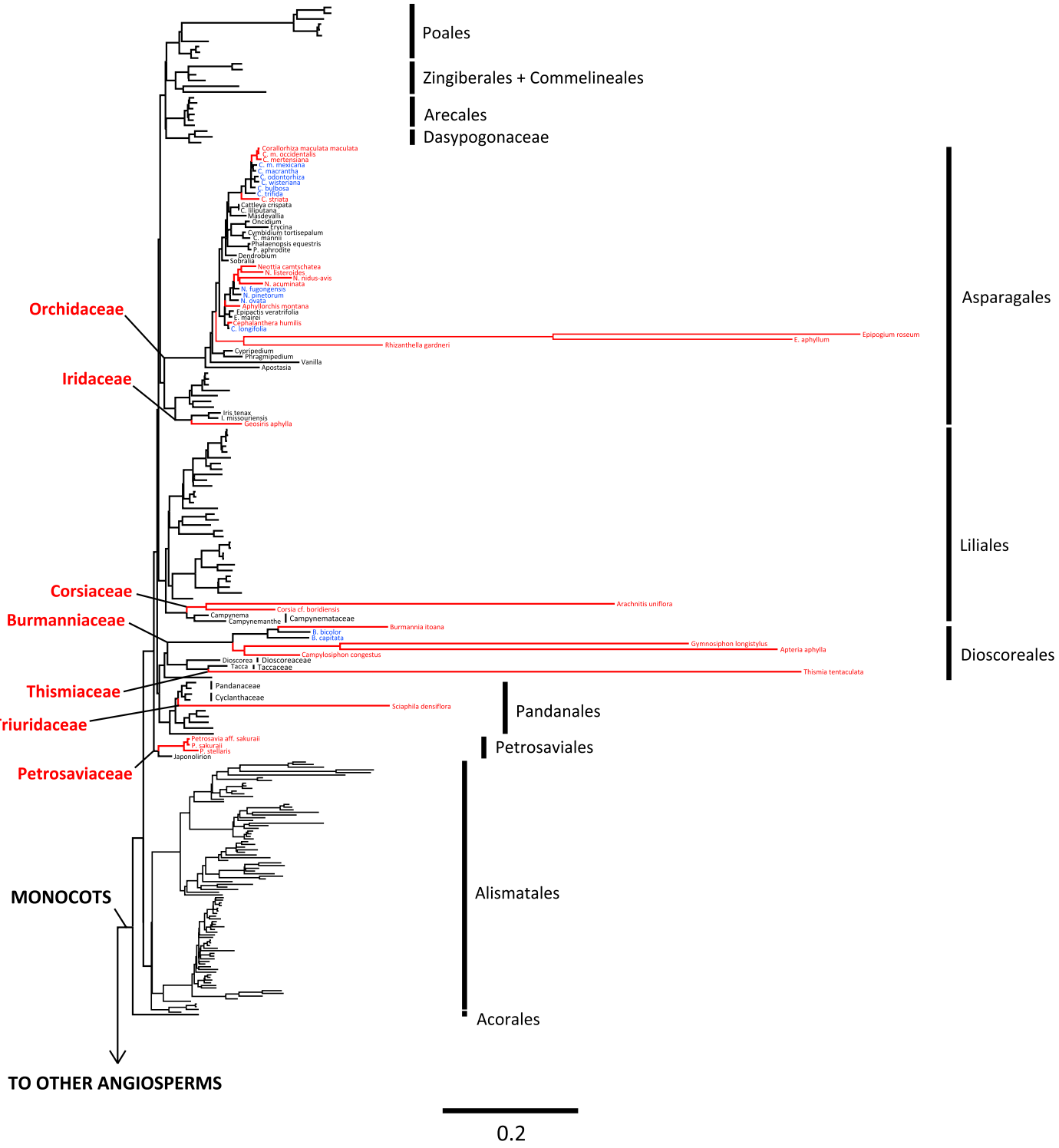


FIGURE 2. Phylogenetic placements of monocot mycoheterotrophic lineages (continuation of likelihood tree in Fig. 1; support values summarized in Figs. 3 and 4 for monocot groups of interest). Red lineages represent fully mycoheterotrophic taxa (with Dollo parsimony interpretations of photosynthesis loss; re-gain not possible once lost); terminals with blue labels are suspected or known partial mycoheterotrophs (*Burmannia capitata*, Burmanniaceae, may be fully autotrophic; Merckx et al., 2010), predicted to retain photosynthetic ability; a few taxa noted in black are photosynthetic plants (not mycoheterotrophic at maturity). Major monocot clades are also noted. The scale bar indicates the estimated number of substitutions per site.

the amino-acid analyses, either as the sister group of the orchid genus *Epipogium*, with the resulting clade then sister to *Arachnitis* in Corsiaceae (for the unpartitioned amino-acid analyses, Appendix S8), or as sister to *Apteria* in Burmanniaceae (for the partitioned amino-acid analyses, Appendix S7). These two odd and conflicting arrangements are, however, very poorly supported in bootstrap

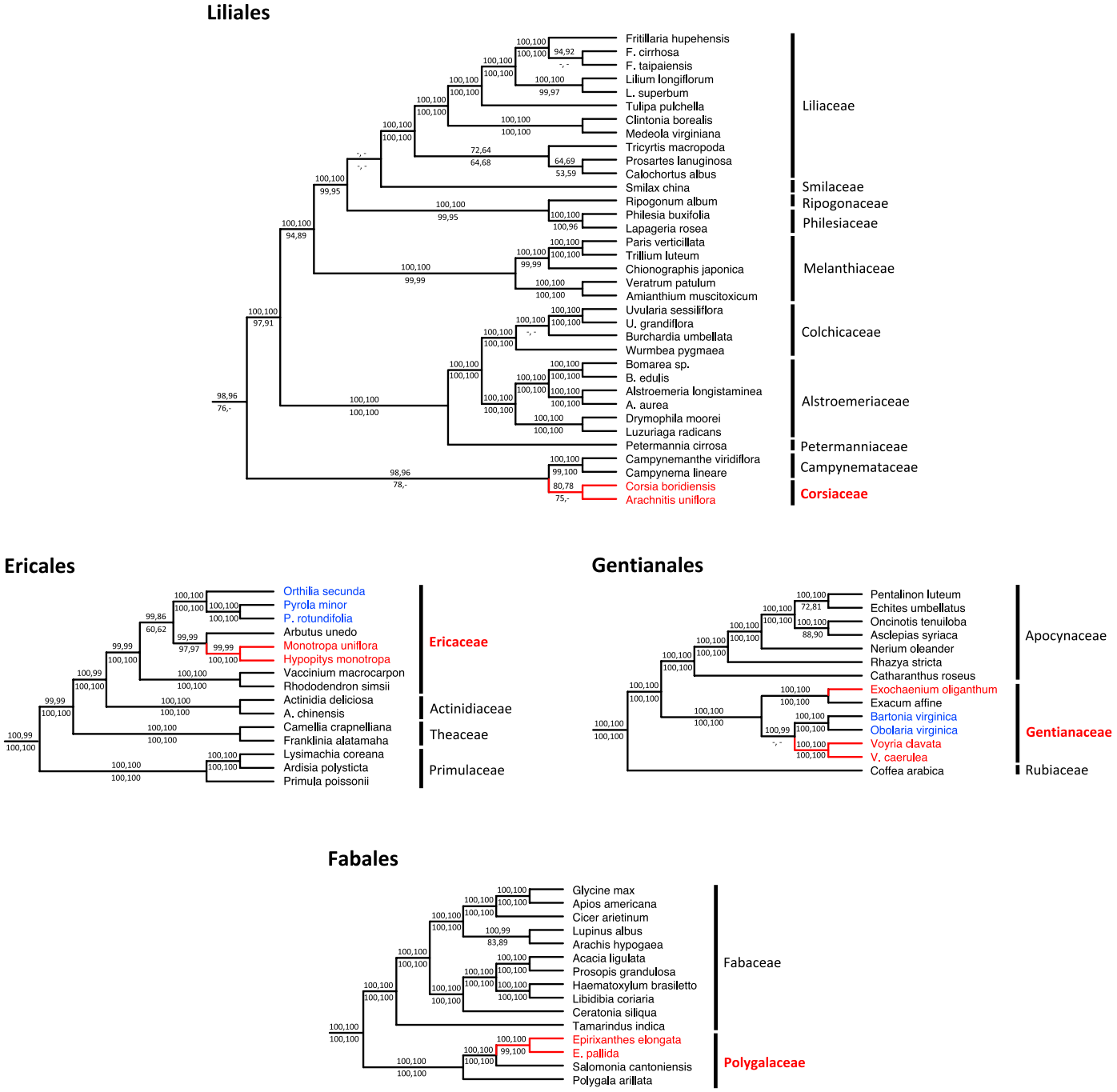


FIGURE 3. Summary of bootstrap support for mycoheterotrophic lineages in three eudicot orders and one monocot order based on likelihood analysis of plastid genome data (DNA sequence data for 77 of the 79 protein-coding genes in most photosynthetic taxa; fewer genes in most mycoheterotrophs, Appendix S3). Tree topology from Figs. 1 and 2. Values above branches indicate the likelihood-based bootstrap support for "gene by codon" partitioned versus unpartitioned likelihood analyses of the DNA sequence data, respectively; values below branches indicate bootstrap support for gene partitioned versus unpartitioned likelihood analyses of the amino-acid matrix, respectively. Branch coloring as in Fig. 1.

analysis (Appendices S7, S8). We tested what happens when either *Epipogium* or *Thismia* is deleted from amino-acid likelihood analysis, using the unpartitioned case. When either is deleted, the other taxon shifts back to a placement that matches that seen in the DNA-based analyses: *Thismia* as sister to *Tacca*, and *Epipogium* as sister to *Rhizanthella*, although these placements are still poorly supported in the corresponding bootstrap analyses (Appendices S9, S10). In the DNA-based likelihood analyses, Burmanniaceae are the sister group of a clade comprising Dioscoreaceae, Taccaceae and Thismiaceae, with 88–91% bootstrap support (Fig. 4; Appendices S5, S6); the monophyly of Burmanniaceae is also strongly supported in these analyses (97–99% bootstrap support). In general, branches within Burmanniaceae are strongly supported in the DNA-based likelihood analyses (Fig. 4; Appendices S5, S6), *Apteria*

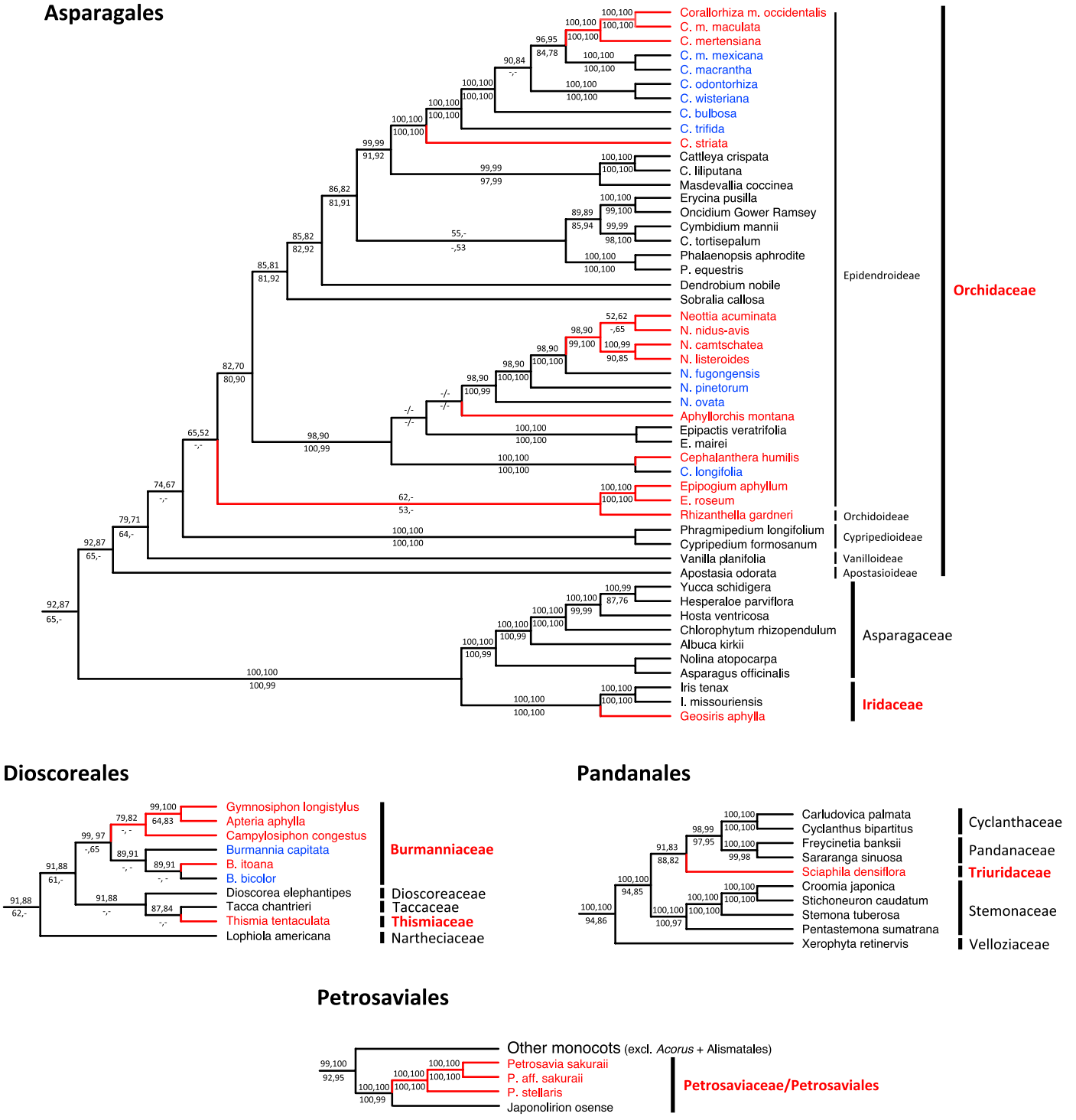


FIGURE 4. Summary of bootstrap support for individual mycoheterotrophic lineages in four monocot orders based on likelihood analysis of plastid genome data (DNA sequence data for 77 of the 79 protein-coding genes in most photosynthetic taxa; fewer genes in most mycoheterotrophs, Appendix S3). Tree topology from Fig. 2. Values above branches indicate the likelihood-based bootstrap support for "gene by codon" partitioned versus unpartitioned likelihood analyses of the DNA sequence data, respectively; values below branches indicate bootstrap support for gene partitioned versus unpartitioned likelihood analyses of the amino-acid matrix, respectively. Branch coloring as in Fig. 1.

and *Gymnosiphon* are well-supported sister taxa (99–100% bootstrap support), with *Campylosiphon* as sister group of this clade (79–82% bootstrap support). *Burmannia itoana* and *B. bicolor* are the sister group of *B. capitata* (89–91% bootstrap support). Alternative arrangements are observed within Burmanniaceae for the partitioned and unpartitioned amino-acid analyses, but they are all very poorly supported (Appendices S7, S8), except for *Apteria-Gymnosiphon* in the unpartitioned amino acid analysis (83%

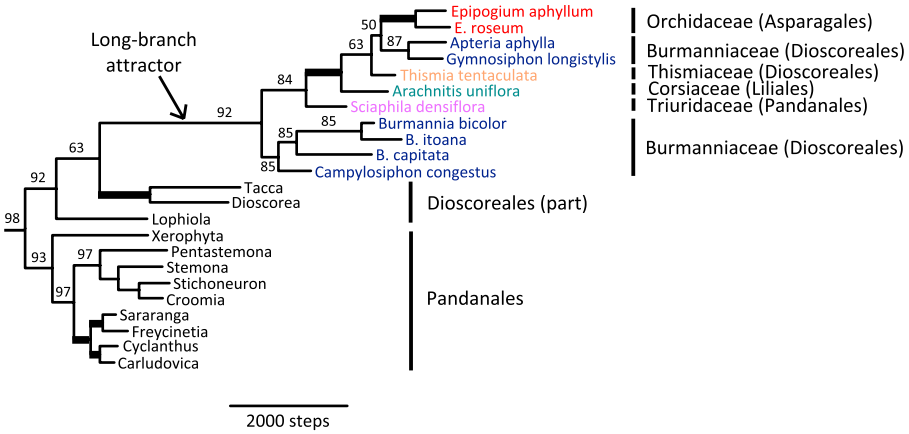


FIGURE 5. A "long-branch attractor" clade comprising multiple rapidly evolving mycoheterotrophic lineages from a parsimony analysis of plastid genome data (DNA sequence data for 77 of the 79 protein-coding genes in most photosynthetic taxa; fewer genes in most mycoheterotrophs, Appendix S3). This figure is a portion of the full analysis (Appendix S11). Mycoheterotrophic taxa from different families are noted in different colors. Bootstrap values are noted on branches (thick branches represent 100% bootstrap support). The scale bar indicates the number of steps.

bootstrap support; Fig. 4). The monophyly of the family and its precise position within Dioscoreales are also poorly supported for both amino-acid analyses (Appendices S7, S8).

In Liliales, *Arachnitis* and *Corsia* are inferred to be sister taxa in Corsiaceae with moderate support in the DNA-based likelihood analyses (78–80% support; Fig. 3; Appendices S5, S6), which also strongly support the family as the sister group of Campynemataceae (*Campynema* and *Campynemanthe*) (96–98% support). The partitioned amino-acid analysis has the same arrangements, with moderate support (75% and 78% for the two relevant branches; Fig. 3, Appendix S7). An unusual placement noted above for the unpartitioned amino-acid analysis, with *Thismia* and the orchid *Epipogium* recovered as sister to *Arachnitis*, and all three sister to *Corsia*, is poorly supported (Fig. 3; Appendix S8).

In Asparagales, Iridaceae (which include *Geosiris*) are inferred to be the sister group of all sampled Asparagales except Orchidaceae. *Geosiris* is the sister group of two sampled Iridaceae here (Figs. 2, 4; Appendices S5–S8). Orchids are inferred to be the sister group of all other Asparagales, with moderate to strong support in DNA-based likelihood analysis (Fig. 4; Appendices S5, S6). Placements of mycoheterotrophic orchids within the family are largely consistent across likelihood analyses (Fig. 4, Appendices S5–S8). Some are poorly supported, including the placement of *Neottia*, *Aphyllorchis*, *Cephalanthera*, and *Epipactis* with respect to each other (short branches connect these genera, and three of the four include fully mycoheterotrophic species; Fig. 2), and a purported sister-group relationship between *Epipogium* and *Rhizanthella* (the two longest-branch orchids; Fig. 2) in three of the four main likelihood analyses (Appendices S5–S7). Minor variations are also observed elsewhere among the analyses, such as the precise position of *C. bulbosa* in *Corallorhiza* (Appendices S5–S8).

Parsimony-based analysis of mycoheterotrophic lineages

The phylogeny inferred from the parsimony analysis groups a clade comprising two orchid species in *Epipogium* as the sister group of two genera in Burmanniaceae, *Apteria* and *Gymnosiphon*; *Thismia* (Thismiaceae), *Arachnitis* (Corsiaceae), and *Sciaphila* (Triuridaceae) are respectively deeper sister groups of this clade, and another portion of Burmanniaceae comprising *Burmannia bicolor*, *B. itoana*, *B. capitata*, and *Campylosiphon* is sister to the remainder of this clade (Fig. 5; Appendix S11). This entire clade is strongly supported, and most of the structure in it is also moderately to strongly supported, presumably reflecting strong parsimony misinference due to long-branch attraction among rapidly-evolving lineages from different families of mycoheterotrophs (see Fig. 2). The parsimony-based placements of the remaining lineages that include full mycoheterotrophs (i.e., Petrosaviaceae; *Geosiris* in Iridaceae; *Corsia* in Liliales; other orchids) (Appendix S11) are otherwise similar to those inferred in the likelihood analyses (Appendices S5–S8).

DISCUSSION

Phylogenomic inference using rapidly evolving and incomplete plastomes

Plastid genomes of mycoheterotrophs have only recently been used in phylogenetic inference: it was once assumed that too many genes (or the entire genome) would be lost, or that retained genes would be too rapidly evolving to be useful (e.g., Cronquist, 1988, p. 467; Merckx et al., 2009). However, multiple recent studies have demonstrated that even highly reduced and rapidly evolving plastid genomes can permit inference of phylogenetic relationships for mycoheterotrophs (e.g., Logacheva et al., 2014; Lam et al., 2015; Mennes et al., 2015a), and that these inferences can be well supported and consistent with other studies based on mitochondrial or nuclear genes. Our simultaneous analysis of mycoheterotrophs from the 10 angiosperm families in which they are found confirms and expands upon this basic result: likelihood analyses can reliably place distantly related mycoheterotrophs in simultaneous analysis, even when their plastid genomes are rapidly evolving and reduced.

Here we investigated a range of DNA- and amino-acid-based substitution models, including partitioned and unpartitioned analyses. Partitioned likelihood analysis takes into account different substitution models or model details that may affect different data subsets; unpartitioned analysis ignores these differences. In general, partitioned and unpartitioned analyses behaved similarly. Overall, we inferred generally well-supported placements for mycoheterotrophic lineages, despite considerable rate elevation in many of

them (Figs. 1, 2) and limited recovery of genes across some taxa (for example, *Arachnitis*, *Sciaphila*, and *Thismia* have only 16, 18, and 7 retained protein-coding genes included here, respectively; Appendix S3). Our four main likelihood analyses are generally congruent with other recent studies, with some exceptions for the two amino-acid-based analyses, discussed below. However, parsimony analysis appears to be strongly misled by long-branch attraction involving distantly related monocot lineages.

Congruence of likelihood analyses with previously published studies

The placements of Ericaceae, Gentianaceae, and Polygalaceae in eudicot phylogeny agree with other studies (e.g., Soltis et al., 2011). Within Ericaceae, two inferred clades—*Pyrola-Orthilia*, representing the pyroloids, and [*Arbutus*, (*Monotropa-Hypopitys*)] representing the arbutoids and monotropoids—are consistent with other studies (e.g., Kron et al., 2002; Braukmann and Stefanović, 2012; Freudenstein et al., 2016; Braukmann et al., 2017). We found variable arrangements of these lineages to each other and to the other sampled photosynthetic lineages in Ericaceae (i.e., *Rhododendron* and *Vaccinium*, representing core Ericaceae; Appendices S5–S8) in the likelihood trees, consistent with continuing uncertainty over several deep relationships in Ericaceae (Freudenstein et al., 2016; Lallemand et al., 2016). Although we did not include the fully mycoheterotrophic member of the pyroloids here (*Pyrola aphylla*), we sampled two photosynthetic members for the genus. A strongly supported sister-group relationship between arbutoids (all photosynthetic) and monotropoids (all mycoheterotrophic), with either observed relationship of this clade to pyroloids and core Ericaceae (Appendices S5–S8), is consistent with two predicted losses of photosynthesis in Ericaceae. There may also be an additional loss of photosynthesis in the family, depending on the phylogenetic position of tribe Pterosporeae (Merckx et al., 2013c), not sampled here. In Polygalaceae, all likelihood analyses unequivocally place *Epirixanthes* as sister to *Salomonia*, with these in turn sister to *Polygala* (Fig. 3; Appendices S5–S8), congruent with the results of Bello et al. (2010) and Mennes et al. (2015b). In Gentianaceae, we inferred three sister pairs in all analyses (*Exochaenium-Exacum*, *Bartonia-Obolaria*, and the two sampled *Voyria* species), but the relative arrangement of these taxa, and the support for these arrangements, varied across analyses (Fig. 3; Appendices S5–S8). *Exochaenium* and *Exacum* were strongly supported as sister taxa in Merckx et al. (2013b), but their study did not include *Bartonia* and *Obolaria*.

For mycoheterotrophic monocots, our inferred placements of *Geosiris* and *Petrosavia*, respectively, as members of Iridaceae (Asparagales) and Petrosaviales (Figs. 2, 4), are consistent with previous studies (Fay et al., 2000; Fuse and Tamura, 2000; Reeves et al., 2001; Davis et al., 2004; Chase et al., 2006; Goldblatt et al., 2008; Lam et al., 2016). Our inferred placements of mycoheterotrophic orchids are consistent with other studies, in particular with Givnish et al. (2015), from which many of the orchid plastomes were taken. One unusual feature is that *Epipogium*, a member of subfamily Epidendroideae not included in Givnish et al. (2015) is weakly supported here as the sister group of *Rhizanthella* in three of four likelihood analyses. *Rhizanthella* is the only included representative of subfamily Orchidoideae, and this arrangement disrupts the monophyly of Epidendroideae. However, support for this sister-group relationship is very poor across all analyses (Fig. 4; Appendices S5–S7) and these two taxa are some of the longest branches in our trees, suggesting that this may be a weak long-branch artifact in likelihood analyses. The odd placement of *Epipogium* in the unpartitioned amino-acid likelihood analysis (Appendix S8) with both *Thismia* and *Arachnitis*, appears to be a weak long-branch artifact, as this arrangement is poorly supported in bootstrap analysis, and *Epipogium* reverts to the orchid clade when *Thismia* is deleted from analysis (Appendix S10). Within *Corallorhiza*, our findings mirror Barrett et al. (2014) where these data came from, although we observed differences between likelihood analyses concerning the placement of *C. bulbosa*, either as sister to *C. odontorhiza* and *C. wisteriana* (weakly supported in the amino-acid analyses, see Appendices S7, S8), or as the sister group of all other species in the genus except *C. striata* and *C. trifida* (moderately to strongly supported in the DNA-based likelihood analyses; Fig. 4; Appendices S5, S6). Within *Neottia*, our findings agree with those of Feng et al. (2016). We also found several arrangements of *Neottia* relative to other closely related orchids (*Aphyllorchis*, *Cephalanthera*, and *Epipactis*); these contrasting positions were not strongly supported in the bootstrap analyses (Fig. 4; Appendices S5–S8).

Mennes et al. (2015a) resolved Corsiaceae as the sister group of Campynemataceae (Liliales) with 100% likelihood-based bootstrap support, using the plastid genomes for these taxa included here, and a variety of combined mitochondrial and nuclear gene data sets. We recovered the same placement in most of our likelihood-based analyses with strong bootstrap support, using an increased plastome sampling in Liliales here (Fig. 3; Appendices S5–S8; see also Givnish et al., 2016). Other studies have found *Arachnitis* to be weakly supported as sister to the rest of Liliales (e.g., Davis et al., 2004; Fay et al., 2006; Petersen et al., 2013), although some of these studies had poor outgroup sampling. A study based on the nuclear 26S rDNA locus recovered Corsiaceae as polyphyletic, with *Corsia* inferred to be the sister group of *Campynema* (Liliales), and *Arachnitis* recovered as the sister group of *Thismia* (Thismiaceae), embedded in a clade otherwise comprising members of Burmanniaceae and Thismiaceae (Neyland and Hennigan, 2003). This earlier finding for Corsiaceae was likely a function of long-branch attraction resulting from elevated substitution rates, limited taxon sampling, and the use of parsimony. It mirrors the strongly supported long-branch attraction artifact that we infer for parsimony analysis here (Fig. 5; Appendix 11), and the more subtle long-branch artifact that we infer for one of the amino-acid likelihood analyses (Appendix S8). The latter is weakly supported in bootstrap analysis and disappears when either of the two longest-branch monocots is excluded (Appendices S9, S10).

The family Burmanniaceae, broadly construed, has been treated as two distinct subtribes, Burmannieae and Thismieae (e.g., Jonker, 1938), or as separate families, Burmanniaceae and Thismiaceae (e.g., Dahlgren and Bremer, 1985; APG, 1998). Modern treatments (APG 2003, 2009, 2016) combine Burmanniaceae and Thismiaceae as one family, Burmanniaceae s.l., based on two studies that included problematic samples (Caddick et al., 2000, 2002; see Lam et al., 2016). In contrast, studies that considered nuclear and mitochondrial data (Merckx et al., 2006, 2009) placed Thismiaceae as the sister group of *Tacca* (Taccaceae) and Burmanniaceae as the sister group of *Dioscorea* (Dioscoreaceae), supporting recognition of Burmanniaceae and Thismiaceae as separate families. Analyses based on one to three plastid genes (Lam et al., 2016) also suggest that Burmanniaceae and Thismiaceae are not each other's closest relatives. Although the latter inferences were very poorly supported, their main tree depicts the same relationships among the families observed here (cf. Fig. 4 here, fig. 2 of Lam et al., 2016). The relationships inferred within

Burmanniaceae in DNA-based likelihood analyses here (Fig. 4; Appendices S5, S6) also agree with those of Merckx et al. (2008). In contrast, the amino-acid likelihood analyses depict different, mostly conflicting arrangements of taxa in the family, and these conflicts are all poorly supported. Oddly, the shortest likelihood tree for the partitioned amino-acid analysis depicts *Thismia* in Burmanniaceae, as sister to *Apteria* (Appendix S7). As noted previously, this poorly supported arrangement is presumably a weak long-branch artifact, because deletion of *Epipogium* reverts *Thismia* to a sister-group relationship with Taccaceae (Appendix S9).

Triuridaceae have sometimes been placed in their own order, Triuridales, based on their morphological distinctiveness (e.g., Dahlgren and Clifford, 1982; Maas-van de Kamer and Weustenfeld, 1998). Their first clear phylogenetic placement came from Chase et al. (2000), who recovered *Sciaphila* (Triuridaceae) as closely related to Cyclanthaceae and Pandanaceae (Pandanales) in a combined analysis of several plastid genes and 18S rDNA (only the latter gene was recovered for *Sciaphila*). A phylogenetic analysis based on 39 morphological characters placed Triuridaceae within Stemonaceae, also in Pandanales (Rudall and Bateman, 2006). More recently, Mennes et al. (2013) used mitochondrial and nuclear evidence to place Triuridaceae as the sister group of a clade comprising Pandanaceae, Cyclanthaceae, and Stemonaceae, with low support. Lam et al. (2015) used the same sampling of plastomes employed here to place *Sciaphila* as the sister group of Cyclanthaceae-Pandanales in likelihood analysis, with strong support across different likelihood analyses, confirming the placement by Chase et al. (2000) of the family based on 18S rDNA. We inferred the same relationship here with slightly reduced support (Fig. 4; Appendices S5–S8), perhaps because we used one fewer plastid gene here (the large *ycf2* reading frame was included by Lam et al., 2015, but excluded here because of alignment difficulties across angiosperms), and support may also have been influenced by the inclusion of many other rapidly evolving mycoheterotrophs. Nonetheless, our inferred placement of Triuridaceae as the sister group of Cyclanthaceae-Pandanales can be considered to be robust, as it is consistent across very different likelihood frameworks (i.e., partitioned and unpartitioned likelihood analysis for DNA and amino-acid data here; codon-based analysis by Lam et al., 2015), and is consistently well supported in all likelihood-based bootstrap analysis.

Propensity for long-branch attraction of different phylogenetic inference approaches

The DNA-based likelihood analyses appear to be insensitive to whether complex or simple data partitioning schemes are employed (see support values in Figs. 3, 4), and our combined analyses with multiple long-branch taxa are consistent with other well-sampled published studies, as noted above. This suggests minimal opportunity for misinference due to long-branch attraction in our two DNA-based likelihood analyses. However, both amino-acid likelihood analyses have poorer support for placements of some lineages (e.g., the relative arrangement of *Voyria* to other taxa in Gentianaceae; Fig. 3; Appendices S7, 8), or have some unusual but poorly supported placements, that conflict between the two amino-acid analyses (i.e., *Thismia* in the partitioned amino-acid analysis, which appears to attract to a long branch in Burmanniaceae, Appendix S7; *Thismia* and the orchid genus *Epipogium* in the unpartitioned amino-acid analysis, which appear to attract to a long branch in Corsiaceae, Appendix S8). Inferred placements of other mycoheterotrophs in the shortest trees recovered in the amino-acid likelihood analyses are otherwise consistent with those seen in the DNA-based likelihood analyses (Appendices S5–S8) and other published studies based on other sources of evidence (see previous section). For the affected taxa, it is possible that a combination of fewer overall characters in the amino-acid data sets (after translation, compared to the longer DNA alignments), leads to reduced bootstrap support for some lineages (especially those that are rapidly evolving and represented by a few genes) and amplifies a weak tendency for several of the longest branches to attract in the amino-acid analyses. It is not clear how to tease apart the possible effect of faster vs. fewer genes on bootstrap support for affected lineages. Nonetheless, the presumed misplacements are limited and poorly supported, and when one of the two most problematic (and longest-branch taxa) is deleted (*Thismia* or *Epipogium*), the other taxon returns to the arrangement seen in the DNA-based likelihood analyses (Appendices S9, S10). This supports the idea that these oddities in amino-acid likelihood analyses represent a mild and surmountable long-branch problem.

Misinference due to long-branch attraction appears to be substantially more problematic for parsimony analysis, which pulls together all members of Burmanniaceae (including photosynthetic members), *Arachnitis* (Corsiaceae: Liliales), *Epipogium* (Orchidaceae: Asparagales), and *Sciaphila* (Triuridaceae: Pandanales) into a single mixed clade with strong bootstrap support. This entire clade is inferred to be the sister group of a subset of Dioscoreales (Fig. 5, Appendix S11). Notably, it includes most of the longest-branch monocot lineages in the likelihood analyses. Because the composition of this clade clashes strongly with our current understanding of monocot systematics (e.g., APG, 2016) and includes some of the longest-branch lineages predicted in our phylogeny (Fig. 2), we infer this very odd result to be a consequence of strong long-branch attraction for parsimony (Felsenstein, 1978, Hendy and Penny, 1989). Consistent with this hypothesis, other more slowly evolving lineages in some of these families (*Corsia*, Corsiaceae; multiple lineages of Orchidaceae) place in positions consistent with the likelihood results here (compare Appendix S11 to Figs. 1–4 and Appendices S5–S8) and with other published studies. Similar long-branch effects likely explain oddly placed lineages in parsimony analyses of mycoheterotrophs that employed only a few genes (e.g., Neyland and Hennigan, 2003; Lam et al., 2016). We did not perform a formal rate analysis of the long-branch lineages here, but base our inferences of elevated rates on simple visual inspections of branch lengths relative to close green relatives in the same or related orders (Figs. 1, 2) (see Lam et al., 2016, for a rate analysis of a subset of the genes employed here).

CONCLUSIONS AND FUTURE DIRECTIONS

Our sampling includes representatives of all 10 families of angiosperms that include fully mycoheterotrophic lineages, but represents only a fraction of the species and the independent losses of photosynthesis that they represent. Notably, our results are largely consistent across very diverse likelihood frameworks (i.e., unpartitioned and partitioned analyses, performed at the level of both DNA and amino-acid sequences), and are generally well supported. Future work should focus on adding additional mycoheterotrophic lineages, which we predict should be straightforward to analyze using model-based phylogenomic inference. Despite our relatively limited

taxon samplings in individual families, the relationships we infer in Burmanniaceae, Ericaceae, Gentianaceae, and Orchidaceae using likelihood inference are consistent with multiple losses of photosynthesis in each case (e.g., see lineages highlighted in red in Figs. 1–4). Future studies should include representatives of additional fully mycoheterotrophic lineages in Ericaceae (*Pyrola aphylla*; see Merckx et al., 2013a), Gentianaceae (*Voyriella* and some species of *Exacum*; Merckx et al., 2013a, b) and others. The greatest number of unsampled lineages of full mycoheterotrophs are in Burmanniaceae (Merckx et al., 2006, 2013c) and the orchids (Merckx et al., 2013a, c); adding additional orchid lineages (both autotrophic and mycoheterotrophic) should also help break up several long branches that may have led to the unusual and poorly supported placements of two mycoheterotrophic orchids observed in likelihood analysis here (*Epipogium* and *Rhizanthella*), and several odd but poorly supported relationships in amino-acid analysis that appear to be weak long-branch artifacts affecting rapidly evolving taxa that retain only a handful of genes. Additional partial mycoheterotrophs would also be valuable to include. Our results also support including rapidly evolving parasitic plant lineages with reduced genomes (e.g., *Cytinus hypocistis*, Cytinaceae, Roquet et al., 2016; *Hydnora visseri*, Hydnoraceae, Naumann et al., 2016; *Cynomorium coccineum*, Cynomoriaceae, Bellot et al., 2016; *Pilostyles* spp., Apodanthaceae, Bellot and Renner, 2016) in similar large-scale plastid phylogenomic analysis (for example, in combined analyses that include all heterotrophic plant lineages). It would also be useful to extend phylogenomic inference to include mitochondrial and nuclear data sets from all major heterotrophic lineages.

Our study demonstrates that mycoheterotrophic lineages can in general be readily included in broadly sampled phylogenomic studies of angiosperm phylogeny using model-based approaches, despite extremely long branches in some mycoheterotrophic lineages and often radical genome reduction. It is the first plastome-based study to include all currently known mycoheterotrophic families and thus provides an excellent framework for studying additional independent origins of mycoheterotrophy in the angiosperms and their molecular evolution, including models of gene loss (e.g., Barrett and Davis, 2012; Graham et al., 2017), and studies of genome structure evolution and changes in selective regimes associated with the origin of heterotrophy in photosynthetic lineages (e.g., Barrett and Davis, 2012; Lam et al. 2015; Schelkunov et al., 2015). More broadly sampled phylogenetic trees of mycoheterotrophs will also be useful for testing hypotheses related to the origin of mycoheterotrophs, such as the evolution of dust seeds (e.g., Eriksson and Kainulainen, 2011), and a predicted association between mycoheterotrophy and root systems with star-like organization, voluminous primary root cortex parenchyma, and complex mycorrhizal colonization patterns (Imhof, 2010). More generally, they will be useful for unraveling the pathways from autotrophy to partial and full mycoheterotrophy (e.g., Merckx et al., 2013c; Lallemand et al., 2016).

ACKNOWLEDGEMENTS

The authors thank Thassilo Franke, Stephanie Lyon, Daniel McNair, Ray Neyland, Thomas W. A. Braukmann, Saša Stefanović, Robert I. Bertin, and other collectors and Botanical Gardens for assistance in obtaining plant material. Several anonymous reviewers provided helpful comments that improved the manuscript. This work was supported by an NSERC Postgraduate Fellowship and UBC Four-Year Fellowship to V.K.Y.L. and an NSERC (Natural Sciences and Engineering Research Council of Canada) Discovery Grant to S.W.G.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

LITERATURE CITED

Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403–410.

APG [Angiosperm Phylogeny Group]. 1998. An ordinal classification for the families of flowering plants. *Annals of the Missouri Botanical Garden* 4: 531–553.

APG. 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. *Botanical Journal of the Linnean Society* 141: 399–436.

APG. 2009. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III. *Botanical Journal of the Linnean Society* 161: 105–121.

APG. 2016. An update of the Angiosperm Phylogeny Group classification for the orders and families of flower plants: APG IV. *Botanical Journal of the Linnean Society* 181: 1–20.

Barkman, T. J., S.-H. Kim, K. M. Salleh, and J. Nais. 2004. Mitochondrial DNA sequences reveal the photosynthetic relatives of *Rafflesia*, the world's largest flower. *Proceedings of the National Academy of Science, USA* 101: 787–792.

Barrett, C. F., W. J. Baker, J. R. Comer, J. G. Conran, S. C. Lahmeyer, J. H. Leebens-Mack, J. Li, G. S. Lim, et al. 2016. Plastid genomes reveal support for deep phylogenetic relationships and extensive rate variation among palm and other commelinid monocots. *New Phytologist* 209: 855–870.

Barrett, C. F., and J. I. Davis. 2012. The plastid genome of mycoheterotrophic *Corallorhiza striata* (Orchidaceae) is in the relatively early stages of degradation. *American Journal of Botany* 99: 1513–1523.

Barrett, C. F., J. V. Freudenstein, J. Li, D. R. Mayfield-Jones, L. Perez, J. C. Pires, and C. Santos. 2014. Investigating the path of plastid genome degradation in an early-transitional clade of heterotrophic orchids, and implications for heterotrophic angiosperms. *Molecular Biology and Evolution* 31: 3095–3112.

Bello, M. A., P. J. Rudall, and J. A. Hawkins. 2010. Combined phylogenetic analyses reveal interfamilial relationships and patterns of floral evolution in the eudicot order Fabales. *Cladistics* 28: 393–421.

Bellot, S., N. Cusimano, S. Luo, G. Sun, S. Zarre, A. Gröger, E. Temsch, and S. S. Renner. 2016. Assembled plastid and mitochondrial genomes, as well as nuclear genes, place the parasitic family Cynomoriaceae in the Saxifragales. *Genome Biology and Evolution* 8: 2214–2230.

Bellot, S., and S. S. Renner. 2016. The plastomes of two species in the endoparasite genus *Pilostyles* (Apodanthaceae) each retain just five or six possibly functional genes. *Genome Biology and Evolution* 8: 189–201.

Bernard, N. 1909. L'évolution dans la symbiose. Les orchidées et leurs champignons commensaux. *Annales des Sciences Naturelles, Botanique* 9: 1–196.

Bodin, S. S., J. S. Kim, and J.-H. Kim. 2016. Phylogenetic inferences and the evolution of plastid DNA in Campynemataceae and the mycoheterotrophic *Corsia dispar* D.L. Jones & B. Gray (Corsiaceae). *Plant Molecular Biology Reporter* 34: 192–210.

Bolin, J. F., K. U. Tennakoon, M. Bin Abdul Majid, and D. D. Cameron. 2017. Isotopic evidence of partial mycoheterotrophy in *Burmannia coelestis*. *Plant Species Biology* 32: 74–80.

Braukmann, T., and S. Stefanović. 2012. Plastid genome evolution in mycoheterotrophic Ericaceae. *Plant Molecular Biology* 79: 5–20.

Braukmann, T. W. A., M. B. Broe, S. Stefanović, and J. V. Freudenstein. 2017. On the brink: the highly reduced plastomes of nonphotosynthetic Ericaceae. *New Phytologist* 216: 254–266.

Burleigh, J. G., K. W. Hilu, and D. E. Soltis. 2009. Inferring phylogenies with incomplete data sets: a 5-gene, 567-taxon analysis of angiosperms. *BMC Evolutionary Biology* 9: 61.

Caddick, L. R., P. J. Rudall, P. Wilkin, and M. W. Chase. 2000. Yams and their allies: Systematics of Dioscoreales. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: systematics and evolution, 475–487. CSIRO Publishing, Victoria, Australia.

Caddick, L. R., P. J. Rudall, P. Wilkin, T. A. J. Hedderson, and M. W. Chase. 2002. Phylogenetics of Dioscoreales based on analyses of morphological and molecular data. *Botanical Journal of the Linnean Society* 138: 123–144.

Cameron, D. D., and J. F. Bolin. 2010. Isotopic evidence of partial mycoheterotrophy in the Gentianaceae: *Bartonia virginica* and *Obolaria virginica* as case studies. *American Journal of Botany* 97: 1272–1277.

Chase, M. W., M. F. Fay, D. S. Devey, O. Maurin, N. Rønsted, T. J. Davies, Y. Pillon, et al. 2006. Multigene analyses of monocot relationships: a summary. *Aliso* 22: 63–75.

Chase, M. W., D. E. Soltis, P. S. Soltis, P. J. Rudall, M. F. Fay, W. H. Hahn, S. Sullivan, et al. 2000. Higher-level systematics of the monocotyledons: an assessment of current knowledge and a new classification. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: systematics and evolution, 3–16. CSIRO, Melbourne, Australia.

Cronquist, A. 1988. The evolution and classification of flowering plants, 2nd ed. New York Botanical Garden, Bronx, NY, USA.

Davis, J. I., D. W. Stevenson, G. Petersen, O. Seberg, L. M. Campbell, J. V. Freudenstein, D. H. Goldman, et al. 2004. A phylogeny of the monocots, as inferred from *rbcL* and *atpA* sequence variation, and a comparison of methods for calculating jackknife and bootstrap values. *Systematic Botany* 29: 467–510.

Dahlgren, R., and K. Bremer. 1985. Major clades of the angiosperms. *Cladistics* 1: 349–368.

Dahlgren, R. M., and H. T. Clifford. 1982. The monocotyledons: a comparative study. Academic Press, London, UK.

Delannoy, E., S. Fujii, C. C. des Francs-Small, M. Brundrett, and I. Small. 2011. Rampant gene loss in the underground orchid *Rhizanthella gardneri* highlights evolutionary constraints on plastid genomes. *Molecular Biology and Evolution* 28: 2077–2086.

Doyle, J. J., and J. L. Doyle. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* 19: 11–15.

Eriksson, O., and K. Kainulainen. 2011. The evolutionary ecology of dust seeds. *Perspectives in Plant Ecology, Evolution and Systematics* 13: 73–87.

Fay, M. F., M. W. Chase, N. Rønsted, D. S. Devey, Y. Pillon, J. C. Pires, G. Peterson, et al. 2006. Phylogenetics of Liliales. *In* J. T. Columbus, E. A. Friar, J. M. Porter, L. M. Prince, and M. G. Simpson [eds.], Monocots: comparative biology and evolution (excluding Poales), 559–565. Rancho Santa Ana Botanic Garden, Claremont, California, USA.

Fay, M. F., P. J. Rudall, S. Sullivan, K. L. Stobart, A. Y. De Brujin, G. Reeves, F. Qamaruz-Zaman, et al. 2000. Phylogenetic studies of Asparagales based on four plastid DNA regions. *In* K. L. Wilson and D. A. Morrison [eds.], Monocots: systematics and evolution, 360–371. CSIRO Publishing, Collingwood, Australia.

Feng, Y. L., S. Wicke, J.-W. Li, Y. Han, C.-S. Lin, D.-Z. Li, T.-T. Zhou, W.-C. Huang, L.-Q. Huang, and X.-H. Jin. 2016. Lineage-specific reductions of plastid genomes in an orchid tribe with partially and fully mycoheterotrophic species. *Genome Biology and Evolution* 8: 2164–2175.

Felsenstein, J. 1978. Cases in which parsimony or compatibility methods will be positively misleading. *Systematic Zoology* 27: 401–410.

Felstenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* 39: 783–791.

Fishbein, M., S. C. K. Straub, J. Boutte, K. Hansen, R. C. Cronn, A. Liston. 2018. Evolution at the tips: *Asclepias* phylogenomics and new perspectives on leaf surfaces. *American Journal of Botany* https://doi.org/10.1002/ajb2.1062.

Freudenstein, J. V., M. B. Broe, and E. R. Feldenkris. 2016. Phylogenetic relationships at the base of Ericaceae: implications for vegetative and mycorrhizal evolution. *Taxon* 65: 794–804.

Fuse, S., and M. N. Tamura. 2000. A phylogenetic analysis of the plastid *mat*K gene with emphasis on Melanthiaceae sensu lato. *Plant Biology* 2: 415–427.

Gitzendanner, M. A., P. S. Soltis, G. K.-S. Wong, B. R. Ruhfel, and D. E. Soltis. 2018. Plastid phylogenomic analysis of green plants: a billion years of evolutionary history. *American Journal of Botany* https://doi.org/10.1002/ajb2.1048.

Givnish, T. J., M. Ames, J. R. McNeal, M. R. McKain, P. R. Steele, S. W. Graham, J. C. Pires, et al. 2010. Assembling the tree of the monocotyledons: plastome sequence phylogeny and evolution of Poales. *Annals of the Missouri Botanical Garden* 97: 584–616.

Givnish, T. J., D. Spalink, M. Ames, S. P. Lyon, S. J. Hunter, A. Zuluaga, W. J. D. Iles, et al. 2015. Orchid phylogenomics and multiple drivers of their extraordinary diversification. *Proceedings of the Royal Society, B, Biological Sciences* 282: 20151553.

Givnish, T. J., A. Zuluaga, I. Marques, V. K. Y. Lam, M. Soto Gomez, W. J. Iles, M. Ames, et al. 2016. Phylogenomics and historical biogeography of the monocot order Liliales: out of Australia and through Antarctica. *Cladistics* 32: 581–605.

Goldblatt, P., A. Rodriguez, M. P. Powell, T. J. Davies, J. C. Manning, M. Van der Bank, and V. Savolainen. 2008. Iridaceae 'Out of Australasia'? Phylogeny, biogeography, and divergence time based on plastid DNA sequences. *Systematic Botany* 33: 495–508.

Goodwin, S., J. D. McPherson, and W. R. McCombie. 2016. Coming of age: ten years of next-generation sequencing technologies. *Nature Reviews Genetics* 17: 333–351.

Graham, S. W., and W. J. D. Iles. 2009. Different gymnosperm outgroups have (mostly) congruent signal regarding the root of flowering plant phylogeny. *American Journal of Botany* 96: 216–227.

Graham, S. W., V. K. Y. Lam, and V. S. F. T. Merckx. 2017. Plastomes on the edge: the evolutionary breakdown of mycoheterotroph plastid genomes. *New Phytologist* 214: 48–55.

Graham, S. W., P. A. Reeves, A. C. E. Burns, and R. Olmstead. 2000. Microstructural changes in noncoding chloroplast DNA: interpretation, evolution and utility of indels and inversions in basal angiosperm phylogenetic inference. 2000. *International Journal of Plant Sciences* 161: S83–S96.

Gruzdev, E. V., A. V. Mardanov, A. V. Beletsky, E. Z. Kochieva, and N. V. Ravin. 2016. The complete chloroplast genome of parasitic flowering plant *Monotropa hypopitys*: extensive gene losses and size reduction. *Mitochondrial DNA, B* 1: 212–213.

Hendy, M. D., and D. Penny. 1989. A framework for quantitative study of evolutionary trees. *Systematic Zoology* 38: 297–309.

Hynson, N. A., K. Preiss, G. Gebauer, and T. D. Bruns. 2009. Isotopic evidence of full and partial myco-heterotrophy in the plant tribe Pyroleae (Ericaceae). *New Phytologist* 182: 719–726.

Imhof, S. 2010. Are monocots particularly suited to develop mycoheterotrophy? *In* O. Seberg, G. Petersen, A. Barfod, and J. I. Davis [eds.], Diversity, phylogeny, and evolution in the monocotyledons, 11–23. Aarhus University Press, Aarhus, Denmark.

Jiang, W., S.-Y. Chen, H. Wang, D.-Z. Li, and J. J. Wiens. 2014. Should genes with missing data be excluded from phylogenetic analyses? *Molecular Phylogenetics and Evolution* 80: 308–318.

Jonker, F. P. 1938. A monograph of the Burmanniaceae. Kemink en Zoon N. V., Utrecht, Sweden.

Kron, K. A., W. S. Judd, P. F. Stevens, D. M. Crayn, and A. A. Anderberg. 2002. Phylogenetic classification of Ericaceae: molecular and morphological evidence. *Botanical Review* 68: 335–423.

Lallemand, F., M. Gaudeul, J. Lambourdière, Y. Matsuda, Y. Hashimoto, and M.-A. Selosse. 2016. The elusive predisposition to mycoheterotrophy in Ericaceae. *New Phytologist* 212: 314–319.

Lam, V. K. Y., V. S. F. T. Merckx, and S. W. Graham. 2016. A few-gene plastid phylogenetic framework for mycoheterotrophic monocots. *American Journal of Botany* 103: 692–708.

Lam, V. K. Y., M. Soto Gomez, and S. W. Graham. 2015. The highly reduced plastome of mycoheterotrophic *Sciaphila* (Triuridaceae) is colinear with its green relatives and is under strong purifying selection. *Genome Biology and Evolution* 7: 2220–2236.

Lanfear, R., B. Calcott, D. Kainer, C. Mayer, and A. Stamatakis. 2014. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evolutionary Biology* 14: 82.

Lanfear, R., P. B. Frandsen, A. M. Wright, T. Senfeld, and B. Calcott. 2016. PartitionFinder 2: new methods for selecting partitioned models of evolution for molecular and morphological phylogenetic analyses. *Molecular Biology and Evolution* 34: 772–773.

Leake, J. R. 1994. The biology of myco-heterotrophic ('saprophytic') plants. *New Phytologist* 127: 171–216.

Lim, G. S., C. F. Barrett, C.-C. Pang, and J. I. Davis. 2016. Drastic reduction of plastome size in the mycoheterotrophic *Thismia tentaculata* relative to that of its autotrophic relative *Tacca chantrieri*. *American Journal of Botany* 103: 1129–1137.

Logacheva, M. D., M. I. Schelkunov, M. S. Nuraliev, T. H. Samigullin, and A. A. Penin. 2014. The plastid genome of mycoheterotrophic monocot *Petrosavia stellaris* exhibits both gene losses and multiple rearrangements. *Genome Biology and Evolution* 6: 238–246.

Logacheva, M. D., M. I. Scheklunov, and A. A. Penin. 2011. Sequencing and analysis of plastid genome in mycoheterotrophic orchid *Neottia nidus-avis*. *Genome Biology and Evolution* 3: 1296–1303.

Logacheva, M. D., M. I. Schelkunov, V. Y. Shtratnikova, M. V. Matveeva, and A. A. Penin. 2016. Comparative analysis of plastid genomes of non-photosynthetic Eriaceae and their photosynthetic relatives. *Scientific Reports* 6: 30042.

Maas-van de Kamer, H, and T. Weustenfeld. 1998. Triuridaceae. *In* K. Kubitzki [ed.], The families and genera of vascular plants, III, Flowering plants: Monocotyledons, 452–458. Springer, Berlin, Germany.

Mennes, C. B., V. K. Y. Lam, P. J. Rudall, S. P. Lyon, S. W. Graham, E. F. Smets, and V. S. F. T. Merckx. 2015a. Ancient Gondwana break-up explains thedistribution of the mycoheterotrophic family Corsiaceae (Liliales). *Journal of Biogeography* 42: 1123–1136.

Mennes, C. B., M. S. Moerland, M. Rath, E. F. Smets, and V. S. F. T. Merckx. 2015b. Evolution of mycoheterotrophy in Polygalaceae: The case of *Epirixanthes*. *American Journal of Botany* 102: 598–608.

Mennes, C. B., E. F. Smets, S. Moses, and V. S. F. T. Merckx. 2013. New insights in the long-debated evolutionary history of Triuridaceae (Pandanales). *Molecular Phylogenetics and Evolution* 69: 994–1004.

Merckx, V., F. T. Bakker, S. Huysmans, and E. Smets. 2009. Bias and conflict in phylogenetic inference of myco-heterotrophic plants: a case study in Thismiaceae. *Cladistics* 25: 64–77.

Merckx, V., L. W. Chatrou, B. Lemaire, M. N. Sainge, S. Huysmans, and E. F. Smets. 2008. Diversification of myco-heterotrophic angiosperms: evidence from Burmanniaceae. *BMC Evolutionary Biology* 8: 178.

Merckx, V., and J. V. Freudenstein. 2010. Evolution of mycoheterotrophy in plants: a phylogenetic perspective. *New Phytologist* 185: 605–609.

Merckx, V., P. Schols, H. Maas-Van De Kamer, P. Maas, S. Huysmans, and E. Smets. 2006. Phylogeny and evolution of Burmanniaceae (Dioscoreales) based on nuclear and mitochondrial data. *American Journal of Botany* 93: 1684–1698.

Merckx, V., M. Stöckel, A. Fleischmann, T. D. Bruns, and G. Gebauer. 2010. ^{15}N and ^{13}C natural abundance of two mycoheterotrophic and a putatively partially mycoheterotrophic species associated with arbuscular mycorrhizal fungi. *New Phytologist* 188: 590–596.

Merckx, V. S. F. T. 2013. Mycoheterotrophy: An introduction. *In* V. Merckx [ed.], Mycoheterotrophy: The biology of plants living on fungi, 1–18. Springer-Verlag, NY, NY, USA.

Merckx, V. S. F. T., J. V. Freudenstein, J. Kissling, M. J. M. Christenhusz, R. E. Stotler, B. Crandall-Stotler, N. Wickett, et al. 2013a. Taxonomy and classification. *In* A. Merckx [ed.], Mycoheterotrophy: the biology of plants living on fungi, 19–101. Springer-Verlag, NY, NY, USA.

Merckx, V. S. F. T., J. Kissling, H. Hentrich, S. B. Janssens, C. B. Mennes, C. D. Specht, and E. F. Smets. 2013b. Phylogenetic relationships of the mycoheterotrophic genus *Voyria* and the implications for the biogeographic history of Gentianaceae. *American Journal of Botany* 100: 712–721.

Merckx, V. S. F. T., C. B. Mennes, K. G. Peay, and J. Geml. 2013c. Evolution and diversification. *In* V. Merckx[ed.], Mycoheterotrophy: the biology of plants living on fungi, 222–226. Springer-Verlag, NY, NY, USA.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. 2010 Gateway Computing Environments Workshop. *GCE* 2010: 1–8.

Naumann, J., J. P. Der, E. K. Wafula, S. S. Jones, S. T. Wagner, L. A. Honaas, P. E. Ralph, et al. 2016. Detecting and characterizing the highly divergent plastid genome of the nonphotosynthetic parasitic plant *Hydnora visseri* (Hydnoraceae). *Genome Biology and Evolution* 8: 345–363.

Neyland, R., and M. Hennigan. 2003. A phylogenetic analysis of large-subunit (26S) ribosome DNA sequences suggests that the Corsiaceae are polyphyletic. *New Zealand Journal of Botany* 41: 1–11.

Nickrent, D. 2017. The parasitic plant connection. Website http://parasiticplants.siu.edu/Relation-Flowering.html [accessed 29 November 2017].

Nickrent, D. L., A. Blarer, Y.-L. Qiu, R. Vidal-Russell, and F. E. Anderson. 2004. Phylogenetic inference in Rafflesiales: the influence of rate heterogeneity and horizontal gene transfer. *BMC Evolutionary Biology* 4: 40.

Normille, D. 2017. Plant scientists plan massive effort to sequence 10,000 genomes [online]. Website http://www.sciencemag.org/news/2017/07/plant-scientists-plan-massive-effort-sequence-10000-genomes [accessed 9 October 2017].

Petersen, G., O. Seberg, and J. I. Davis. 2013. Phylogeny of the Liliales (monocotyledons) with special emphasis on data partition congruence and RNA editing. *Cladistics* 29: 274–295.

Rai, H. S., H. O'Brien, P. A. Reeves, R. G. Olmstead, and S. W. Graham. 2003. Inference of higher-order relationships in the cycads from a large chloroplast data set. *Molecular Phylogenetics and Evolution* 29: 350–359.

Rambaut, A. 2002. Se-Al v. 2.0a11: Sequence alignment program. Available from http://tree.bio.ed.ac.uk/soft ware/seal/ [accessed 15 January 2016].

Rasmussen, H. N. 1995. Terrestrial orchids: From seed to mycotrophic plant. Cambridge University Press, Cambridge, UK.

Reeves, G., M. W. Chase, P. Goldblatt, P. Rudall, M. F. Fay, A. V. Cox, B. Lejeune, and T. Souza-Chies. 2001. Molecular systematics of Iridaceae: evidence from four plastid DNA regions. *American Journal of Botany* 88: 2074–2087.

Roquet, C., É. Coissac, C. Cruaud, M. Boleda, F. Boyer, A. Alberti, L. Gielly, P. Taberlet, W. Thuiller, J. Van Es, and S. Lavergne. 2016. Understanding the evolution of holoparasitic plants: the complete plastid genome of the holoparasite *Cytinus hypocistis* (Cytinaceae). *Annals of Botany* 118: 885–896.

Ross, T. G., C. F. Barrett, M. Soto Gomez, V. K. Y. Lam, C. L. Henriquez, D. H. Les, J. I. Davis, et al. 2016. Plastid phylogenomics and molecular evolution of Alismatales. *Cladistics* 32: 160–178.

Rudall, P. J., and R. M. Bateman. 2006. Morphological phylogenetic analysis of Pandanales: testing contrasting hypotheses of floral evolution. *Systematic Botany* 31: 223–238.

Ruhfel, B. R., M. A. Gitzendanner, P. S. Soltis, D. E. Soltis, and J. G. Burleigh. 2014. From algae to angiosperms - Inferring the phylogeny of green plants (Viridiplantae) from 360 plastid genomes. *BMC Evolutionary Biology* 14: 23.

Saarela, J. M., and S. W. Graham. 2010. Inference of phylogenetic relationships among the subfamilies of grasses (Poaceae: Poales) using meso-scale exemplar-based sampling of the plastid genome. *Botany* 88: 65–84.

Sanderson, M. J., M. F. Wojciechowski, J.-M. Hu, T. Sher Khan, and S. G. Brady. 2000. Error, bias, and long-branch attraction in data for two chloroplast photosystem genes in seed plants. *Molecular Biology and Evolution* 17: 782–797.

Schelkunov, M. I., V. Y. Shtratnikova, M. S. Nuraliev, M.-A. Selosse, A. A. Penin, and M. D. Logacheva. 2015. Exploring the limits for reduction of plastid genomes: a case study of the mycoheterotrophic orchids *Epipogium aphyllum* and *Epipogium roseum*. *Genome Biology and Evolution* 7: 1179–1191.

Soltis, D. E., S. A. Smith, N. Cellinese, K. J. Wurdack, D. C. Tank, S. F. Brockington, N. F. Refulio-Rodriguez, et al. 2011. Angiosperm phylogeny: 17 genes, 640 taxa. *American Journal of Botany* 98: 704–730.

Stamatakis, A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30: 1312–1313.

Stamatakis, A., P. Hoover, and J. Rougemont. 2008. A rapid bootstrap algorithm for the RAxML web servers. *Systematic Biology* 57: 758–771.

Statts, M., R. H. J. Erkens, B. van der Vossenberg, J. J. Wieringa, K. Kraajieveld, B. Stielow, J. Geml, J. E. Richardson, and F. T. Bakker. 2013. Genomic treasure troves: complete genome sequencing of herbarium and insect museum specimens. *PloS ONE* 8: e69189.

Swofford, D. L. 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4.0b10. Sinauer, Sunderland, MA, USA.

Tedersoo, L., P. Pellet, U. Koljalg, and M. A. Selosse. 2007. Parallel evolutionary paths to mycoheterotrophy in understorey Ericaceae and Orchidaceae: ecological evidence for mixotrophy in Pyroleae. *Oecologia* 151: 206–217.

Wicke, S., K. F. Müller, C. W. de Pamphilis, D. Quandt, S. Bellot, and G. Schneeweiss. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proceedings of the National Academy of Sciences, USA* 113: 9045–9050.

Wickett, N. J., S. Mirarab, N. Nguyen, T. Warnow, E. Carpenter, N. Matasci, S. Ayyampalayam, et al. 2014. Phylotranscriptomic analysis of the origin and early diversification of land plants. *Proceedings of the National Academy of Sciences, USA* 111: E4859–E4868.

Wiens, J. J., and J. Tiu. 2012. Highly incomplete taxa can rescue phylogenetic analyses from the negative impacts of limited taxon sampling. *PloS ONE* 7: e42925.

Wyman, S. K., R. K. Jansen, and J. L. Boore. 2004. Automatic annotation of organellar genomes with DOGMA. *Bioinformatics* 20: 3252–3255.

Yang, Z. 2006. Computational Molecular Evolution. Oxford University Press, Oxford, UK.

Zgurski, J. M., H. S. Rai, Q. M. Fai, D. J. Bogler, J. Francisco-Ortega, and S. W. Graham. 2008. How well do we understand the overall backbone of cycad phylogeny? New insights from a large, multigene plastid data set. *Molecular Phylogenetics and Evolution* 47: 1232–1237.

Zimmer, K., N. A. Hynson, G. Gebauer, E. B. Allen, M. F. Allen, and D. J.Read. 2007. Wide geographical and ecological distribution of nitrogen and carbon gains from fungi in pyroloids and monotropoids (Ericaceae) and in orchids. *New Phytologist* 175: 166–175.